Technical Progress Report 2/1/94 - 4/30/94
Construction of a Connectionist Network Supercomputer
University of California, Berkeley
ONR URI Grant No. N00014-92-J-1617

1 Abstract

We have made progress in several areas this quarter:

- The VLSI implementation of the first Torrent processor, T0, is nearing completion; work also continues on the CNS-1 network interface chip, Hydrant.
- We have made steady progress in the development of support software for the Torrent processor.
- Work continued in CNS-1 performance evaluation and architecture refinement.
- We have worked on adapting speech and vision algorithms for the CNS-1.

The CNS-1 project continues to have a significant effect on the education of graduate and undergraduate students at our institution. There are currently 16 Ph.D., 1 M.S., and 2 B.S. students associated with the project (some are paid through supporting agencies other than the ONR). Also, many of the design principles, VLSI building blocks, and CAD tools developed as part of the implementation of the T0 processor are now used in CS250, Graduate VLSI Systems Design, here at Berkeley.

The CNS-1 project was recently represented at the ARPA AVIS (advanced vision systems) workshop held at Caltech, hosted by Carver Mead, on March 22 and 23. We were invited to give a talk about our work on the CNS-1 system. A goal of the meeting was to bring together vision researchers, commercial and military users of advanced vision systems with hardware designers. Our presentation of the CNS-1 system focused on our T0 processor, now in design. We presented its architecture and its potential for vision and image processing applications. The presentation generated considerable industry interest and we are now following up on the inquires about joint projects or transferring some of the CNS-1 technology. A set of our slides from the meeting are included.

2 Technical Status

2.1 Software and Applications

This quarter saw considerable progress in both high and low level software aspects of the project. The final production phase of the T0 chip has led to a number of new demands

on the design and test support software and these are being fully met. The construction of signal processing libraries for T0 is also moving ahead well. At the higher levels, the Sather 1.0 compiler is now fully translating most programs and the remaining problems will be corrected in the coming quarter. This has enabled us to move forward with the implementation design for parallel Sather 1.0 and some related design and documentation changes. Our studies of mapping large connectionist applications to parallel machines has also advanced well and is moving on to a new class of tasks.

Some new experiments using the T0 processor for low-level vision applications show promising results. Based on our cycle-accurate functional simulator, we have developed several T0 application programs. MPEG decoding has been one test case. With a 160 by 128 pixel frame, the inverse DCT part of the decoding task takes 1.02ms on the T0, compared with 21.07ms on a Sparc2 processor. Another experimental application is part of an image processing task from the Berkeley PATH project. Here the task involves a low-level convolution followed by calculation of a disparity map. The T0 version of the calculations gives an approximate 150 times speedup over the Sparc-10 version.

This quarter we have also begun work on mapping speech decoding (the matching of processed acoustic data to phonemes and words) to the SPERT system (the workstation based system built around the T0 chip.) Speech decoding is a computationally intensive task, the structure of which maps well to T0's vector capability. We plan to implement a version of the decoder with a carefully designed framework that will permit SPERT assembly code to be developed in a modular fashion. It is hoped that different decoding strategies can be plugged in, modified or removed by different users with little trouble, facilitating research on speech decoding. In the future, this same framework will be used for the CNS-1 and its variants. This framework and its initial implementation (using the Viterbi algorithm, a form of dynamic programming) on SPERT should be completed in a few months.

2.2 Hardware Development

Testing of the high-speed interface test chip was completed this quarter. As outlined in previous reports a circuit board containing two of the interface chips (transmitter and receiver) and auxiliary circuits was designed and fabricated. This board incorporates two features to be used on the SPERT board design: Chip-on-board packaging and elastomeric test connectors. Both of these features have proven successful for the interface test board, and operation of the 0.8 micron chips has been verified. This chip mimics many of the features of the Rambus interface to transfer data over short distances at a 250 MHz rate. The chips function as expected, however at 20% lower speed. A revision of the chip has been designed and sent to MOSIS for fabrication. Several of the circuits from these test chips are being used in the CNS-1 high-speed network, currently under design.

2.3 Analog VLSI pre-processors

This quarter has brought the first positive results in our evaluation of analog VLSI auditory pre-processors in speech recognition systems. As outlined in previous reports, a silicon auditory model of spectral shape, with on-chip support for efficient communications and parameter storage, has been designed, fabricated, and tested. Also outlined in previous reports has been a prototype system that connects the analog pre-processor with a commercial speech-recognition software library.

Using this system, we are conducting preliminary experiments using an 200-speaker, isolated-word, telephone-quality speech database. These early experiments do not attempt to exploit the unique characteristics of auditory models; instead, we use the auditory chip as a filterbank, and use conventional techniques for converting filterbank outputs into features suitable for speech recognition. We are comparing it with a software-based front end, PLP (Perceptual Linear Prediction), that has found wide use in military, commercial, and academic speech recognition systems.

Our first experiments show that for a simple two-word recognition task (yes, no), the analog VLSI auditory preprocessor performs as well as the PLP front end (98.75% correct); the chip consumes 5 milliwatts of power (at 5V) while processing this speech database. We are currently investigating techniques for converting the chip outputs into features suitable for speech recognition; better feature extraction will be necessary to achieve acceptable performance on harder speech recognition tasks, such as digit recognition.

Also in this quarter, an article we submitted to the technology magazine IEEE Micro on our analog VLSI auditory pre-processor has been accepted for publication; the article will appear in the June issue.

3 Presentations

John Wawrzynek, "The CNS-1 System," ARPA/AVIS Workshop, Caltech, Pasadena, CA, March 22–23, 1994.

Nelson Morgan, "Connectionist Speech Recognition," ARPA/AVIS Workshop, Caltech, Pasadena, CA, March 22–23, 1994.

4 Publications

Lazzaro, J., Wawrzynek, J., and Kramer, A. (1994). "System Technologies for Silicon Auditory Models". IEEE Micro, June (in press).

Mueller, S., and Gomes, B., "A Performance Analysis of CNS-1 on Sparse Connectionist Networks". ICSI Technical Report, TR-94-009, February 1994.